

# *Deceptive minds in digital spaces: the influence of the dark triad on posting fake online reviews*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Borghi, M. ORCID: <https://orcid.org/0000-0002-4150-1595> and Ratcharak, P. (2025) Deceptive minds in digital spaces: the influence of the dark triad on posting fake online reviews. *Psychology & Marketing*, 42 (7). pp. 1932-1945. ISSN 1520-6793 doi: 10.1002/mar.22212 Available at <https://centaur.reading.ac.uk/122139/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/mar.22212>

Publisher: Wiley

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

RESEARCH ARTICLE OPEN ACCESS

# Deceptive Minds in Digital Spaces: The Influence of the Dark Triad on Posting Fake Online Reviews

Matteo Borghi<sup>1</sup>  | Phatcharasiri Ratcharak<sup>2</sup>

<sup>1</sup>Henley Business School, University of Reading, Reading, UK | <sup>2</sup>Greenwich Business School, University of Greenwich (Maritime Campus), London, UK

**Correspondence:** Matteo Borghi ([m.borghi4@henley.ac.uk](mailto:m.borghi4@henley.ac.uk))

**Received:** 2 September 2024 | **Revised:** 9 March 2025 | **Accepted:** 10 March 2025

**Keywords:** Dark Triad | electronic Word of Mouth | fake reviews | Interpersonal Deception Theory | Machiavellianism | narcissism | psychopathy

## ABSTRACT

In the digital age, online reviews play a critical role in consumer decision-making. Despite growing interest in fake online reviews, the personality antecedents of posting such reviews remain largely unexplored. This study addresses this gap by investigating the influence of Dark Triad personality traits on the behavior of posting fake reviews. Leveraging Interpersonal Deception Theory, we apply natural language processing techniques to detect fake reviews in Amazon's data set and employ logistic regression modeling to examine the relationship between these traits and deceptive online behavior. Our findings reveal that Machiavellians use fake reviews as calculated tools to manipulate perceptions, whereas narcissists, driven by their need for validation and fear of detection, avoid posting fake reviews to protect their credibility and online self-image. Furthermore, psychopaths' impulsivity and short-term focus deter them from the cognitively demanding task of crafting plausible fake reviews. These new insights into the distinct effects of each Dark Triad trait on online deceptive behavior enhance theoretical models and offer practical implications. Incorporating personality-based metrics into fake review detection systems can help businesses safeguard their platform integrity and promote authentic consumer feedback.

## 1 | Introduction

The digital era has significantly influenced consumer behavior, making online reviews a vital resource for decision-making regarding products and services (Banerjee and Chua 2023). However, the reliance on user-generated content has also led to deceptive practices, such as the spread of fake reviews, which misrepresent experiences and mislead consumers (Birim et al. 2022). These fraudulent reviews undermine the credibility of online platforms and present substantial challenges for both consumers and businesses (Walther et al. 2023). Consumers often struggle to verify the authenticity of online reviews, while businesses face the potential damage to their reputations from fake feedback (Sahut et al. 2024). With the threat that fake reviews pose to market integrity and consumer trust, major digital platforms hosting online reviews such as Amazon, Trustpilot, and Expedia have begun collectively combating the

spread of fake reviews (Vallance 2023). Governments are also responding with the aim of increasing transparency online and ultimately enabling consumers to make informed decisions based on authentic feedback. For instance, the UK government is set to introduce legislation that classifies fake reviews as a prohibited practice and holds online platforms accountable for the content they host (UK Government 2024).

The increasing prevalence of fake reviews has garnered significant attention across various academic fields, including information systems, engineering, and marketing (Wu et al. 2020). Researchers have primarily focused on enhancing detection methods through advanced data analysis and natural language processing (NLP) techniques (e.g., Birim et al. 2022; Hajek et al. 2023). However, recent syntheses of the fake review literature show that the antecedents of fake reviews remain largely unexplored (Sahut et al. 2024), particularly the

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Psychology & Marketing* published by Wiley Periodicals LLC.

psychological and motivational elements that drive consumers to engage in this behavior (Birim et al. 2022; Wu et al. 2020). Within this research area, psychology scholars suggest that personality traits play a crucial role in one's decision-making to engage in unethical behaviors (Harrison et al. 2018). Existing research indicates that certain personality traits—specifically Machiavellianism, narcissism, and psychopathy, collectively known as the *Dark Triad*, significantly influence unethical activities and antisocial behaviors (Miller et al. 2019). According to Paulhus and Williams (2002), these traits may potentially influence one's process regarding ethical decision-making due to their callous, self-serving, and manipulative characteristics.

Recent research has expanded the investigation of Dark Triad personality traits to the domain of electronic Word of Mouth (eWOM), exploring their influence on online consumer behaviors such as the intention to exaggerate reviews (Kapoor et al. 2021) and their impact on review ratings and perceived helpfulness (Yousaf and Kim 2023). While a growing body of literature identifies the Dark Triad as a significant driver of deceptive behaviors (Miller et al. 2019), prior research in online review contexts (Kapoor et al. 2021) has primarily focused on intentions rather than actual deceptive behaviors. However, as noted by Webb and Sheeran (2006) in their meta-analyses, intentions do not consistently translate into behaviors, a pattern that also applies to deceptive behaviors linked to the Dark Triad (Harrison et al. 2018). Moreover, research highlights that each Dark Triad trait manifests uniquely in individual behavior (Furnham et al. 2013). For instance, Kapoor et al. (2021) found distinct effects of narcissism and psychopathy on the intention to post exaggerated reviews, whereas studies using behavioral data (e.g., Bogolyubova et al. 2018; Jones and Paulhus 2017; Yousaf and Kim 2023) suggest these traits drive different behaviors in digital settings. This underscores the importance of understanding how individuals with Dark Triad traits rationalize and enact deceptive practices like posting fake reviews, given the various ways these traits can manifest in online settings. By addressing this unexplored area, our study seeks to answer the following research question:

### **Do Dark Triad personality traits influence the posting of fake online reviews?**

To address the gap, our study pioneers the use of Interpersonal Deception Theory (IDT) (Buller and Burgoon 1996) to explore the role of Dark Triad traits in the actual behavior of posting fake online reviews. Thus, we extend prior research on deceptive online behaviors by shifting the focus from intentions to deceive, as explained through Moral Disengagement Theory (MDT) (e.g., Kapoor et al. 2021), to actual deceptive behavior, applying IDT to provide a more behaviorally grounded understanding of fake review posting. To achieve this, we leverage a large, publicly available data set of balanced fake and genuine Amazon product reviews (Saxena 2018; Kronrod et al. 2023) and employ NLP techniques to identify language associated with the Dark Triad (Yousaf and Kim 2023). By independently and concurrently analyzing the influence of these traits on deceptive behavior online, our study responds to the call for further research on the antecedents of fake reviews (Sahut et al. 2024), particularly those linked to psychological motives (Birim et al. 2022; Wu et al. 2020). By examining Machiavellianism, narcissism, and psychopathy among online reviewers, the research provides new

insights into their distinct psychological and strategic underpinnings. Machiavellians leverage fake reviews as calculated tools to manipulate perceptions, while narcissists, driven by their need for validation and fear of detection, refrain from deceptive practices to protect their credibility and online self-image. Moreover, psychopaths' impulsivity and short-term focus deter them from the cognitively demanding task of crafting plausible fake reviews. This study advances the understanding of deception in eWOM (e.g., Yousaf and Kim 2023; Zhang et al. 2016), contributing novel perspectives on the interplay between psychological traits and manipulative behaviors in digital contexts. Our findings, therefore, offer practical guidance for managers and platform developers in combating fake reviews. By integrating metrics to detect Dark Triad traits into review filtering systems and tailoring strategies to address deceptive behaviors, businesses can enhance fake review detection accuracy while promoting authentic feedback to build trust and maintain platform integrity.

The article is organized as follows: Section 2 reviews the literature on the Dark Triad, and fake reviews and introduces the main theoretical framework. Following this, the hypotheses linking Machiavellianism, narcissism, and psychopathy to the behavior of posting fake online reviews are presented. Section 3 outlines our research methodology and empirical models while Section 4 presents and interprets the results of our analyses. Finally, Section 5 discusses our findings, highlighting theoretical contributions and practical implications, and concludes with the study's limitations and suggestions for future research.

## **2 | Literature Review**

### **2.1 | The Dark Triad**

The term “Dark Triad,” originally conceived by the seminal study of Paulhus and Williams (2002), brings under the same umbrella three “socially aversive” personality traits, namely, Machiavellianism, narcissism, and psychopathy. These traits were chosen by the authors for the large empirical literature surrounding them. From that point in time, studies in the Dark Triad have flourished, especially in psychology, due to Dark Triad traits' potential impact on numerous realms of human behavior, from professional settings to interpersonal relationships (Lee et al. 2013). Scholars in this area have paid particular attention to how these traits influence antisocial behaviors across various domains (Hollebeek et al. 2022; Miller et al. 2019). Empirical evidence shows how rating high in any of the Dark Triad traits increases the likelihood of individuals engaging in unethical behaviors, including fraud, cheating, and manipulation (Harrison et al. 2018; Jones 2014). These deviant behaviors can naturally lead to negative outcomes for organizations and teams, by affecting trust and cooperation in both personal and professional relationships (Spain et al. 2014). However, as Miller et al. (2019) suggest, to fully comprehend Dark Triad traits' contribution to decision-making processes and ultimately behaviors, there is a need to understand their overlapping and distinctive aspects. Indeed, as per the initial conceptualization of Paulhus and Williams (2002), while each of these traits shares common features such as callousness and a propensity for manipulation, they each exhibit unique

characteristics that influence behavior differently in various contexts (Jones and Paulhus 2017).

Machiavellianism, stemming from the work of Niccolò Machiavelli, is characterized by a cynical view of others, combined with a manipulative interpersonal style and a focus on self-interest (Malär and Giuffredi-Kähr 2024). Those scoring high in Machiavellianism distrust others and may exploit them for personal benefit, often using manipulative tactics (Drouin et al. 2016). Narcissists display grandiosity, dominance, and superiority, often perceiving themselves as superior to others and seeking admiration (Lambert and Desmond 2013). This involves an excessive preoccupation with oneself, including fantasies of success and a sense of entitlement to special treatment (Bowen et al. 2022). Psychopathy, associated with low levels of empathy and anxiety, is defined by high impulsivity and thrill-seeking behaviors (Patrick et al. 2009). With this, individuals characterized by this trait tend to engage in antisocial behaviors, showing little remorse or guilt for their actions and disregarding others' well-being (Karampournioti et al. 2018).

The effect of Dark Triad traits has largely been simplified due to their commonality. Yet, there is an ongoing debate about their structure and organization. For example, some researchers argue that Machiavellianism and psychopathy may be redundant due to their substantial overlap in their measurements (Rogoza et al. 2019). Other scholars suggest the formation of a "Dark Dyad," excluding narcissism based on its distinctiveness from Machiavellianism and psychopathy (Kowalski et al. 2016). For

instance, Jones and Paulhus (2017) found that although both Machiavellianism and psychopathy involve deceit, individuals with high Machiavellianism are more careful in calculating the risk of punishment than those with high psychopathy. Nevertheless, narcissism with its focus on self-admiration and grandiosity is distinct from these two traits, which are antagonistic and manipulative in nature (Back 2018). Table 1 presents a comparison of the Dark Triad traits across various dimensions, highlighting their similarities and differences.

## 2.2 | Fake Reviews

Consumers tend to trust online reviews more than direct promotional content from businesses (Harrison-Walker and Jiang 2023). Yet, this trust is undermined when there is suspicion of fake reviews (Moon et al. 2021). Fake reviews are defined as "deceptive reviews provided with an intention to mislead consumers in their purchase decision-making, often by reviewers with little or no actual experience with the products or services being reviewed" (Zhang et al. 2016, 457). This definition emphasizes the intention to deceive consumers and captures the deliberate nature of fake reviews. They are strategically crafted to manipulate consumer perceptions (Chatterjee et al. 2023). Building on this, the UK Government (UK Department for Business & Trade 2023) further underscores that fake reviews are specifically designed to mislead. Highlighting the extent of the issue, a report from U.S. PIRG suggests that ~30%–40% of online reviews may be fake,

**TABLE 1** | Comparative analysis of Dark Triad traits: characteristics, similarities, and behavioral implications.

Trait	Machiavellianism	Narcissism	Psychopathy
Core characteristics	Cynical, manipulative, self-interested (Malär and Giuffredi-Kähr 2024; Drouin et al. 2016)	Grandiosity, self-focus, need for validation (Morf and Rhodewalt 2001; Bowen et al. 2022)	Emotional callousness, impulsivity, low empathy (Paulhus and Williams 2002; Patrick et al. 2009)
Similarities	All three traits share callousness and a propensity for manipulation (Paulhus and Williams 2002; Lee et al. 2013), and increase the likelihood of unethical behaviors such as fraud and cheating (Harrison et al. 2018; Jones 2014).		
Key distinction	Calculated manipulation, strategic in achieving personal goals (Jonason and Webster 2012)	Need for external admiration and social approval (Besser et al. 2016)	Impulsivity and short-term focus (Patrick et al. 2009)
View of others	Distrustful, manipulative, exploitative (Drouin et al. 2016)	Perceives self as superior, seeks admiration (Bowen et al. 2022)	Low empathy, disregard for others' well-being (Karampournioti et al. 2018)
Focus/ orientation	Self-interest, goal-directed manipulation (Malär and Giuffredi-Kähr 2024)	Self-enhancement, social desirability (Morf and Rhodewalt 2001)	Short-term gratification, thrill-seeking (Harrison et al. 2018; Furnham et al. 2013)
Behavioral tendencies	Deceptive behaviors (e.g., lying and misleading others), strategic manipulation (Trevino and Youngblood 1990)	Self-enhancement behaviors, exploiting others for attention (Emmons 1987; Besser et al. 2016)	Antisocial, impulsive behaviors, minimal remorse (Harrison et al. 2018; Karampournioti et al. 2018)
Social interaction	Manipulates for personal gain, uses tactics like ingratiating, supplication, and self-disclosure (Jonason and Webster 2012)	Seeks admiration, self-promotion, maintains social status (Morf and Rhodewalt 2001; Besser et al. 2016)	Thrill-seeking, risky behaviors, no concern for social norms, short-term actions (Patrick et al. 2009)

indicating the widespread nature of this deceptive practice (Fawler 2023). While fake reviews aim to deceive, it is also important to distinguish them from authentic and exaggerated reviews. Authentic reviews reflect genuine consumer experiences and are primarily posted to communicate real opinions (Hajek et al. 2023). However, reviewers may misrepresent these experiences by overstating certain positive or negative aspects, leading to exaggerated reviews (Kapoor et al. 2021). Unlike fake reviews, which are entirely fabricated, exaggerated reviews remain grounded in actual consumer experiences (Chatterjee et al. 2023; Kapoor et al. 2021). As Banerjee (2022) suggests, exaggeration refers to a writing style that is not necessarily associated with fake reviews.

Fake reviews, compared to authentic ones, have been found to exert greater influence on consumer decision-making (Hajek et al. 2023), which can severely undermine the value of online consumer reviews, such as their perceived authenticity and helpfulness (Lie et al. 2024). Indeed, perceived authenticity significantly influences review helpfulness, shaping whether consumers find a review credible and useful (Lie et al. 2024). Fake reviews often provide conflicting information, causing cognitive dissonance as consumers attempt to reconcile the discrepancies (Chatterjee et al. 2023). This situation highlights the psychological burden on consumers to discern the authenticity of online reviews. The motivations behind producing fake reviews can be various. Intrinsic factors may include personal grudges or brand loyalty, while extrinsic factors often involve monetary rewards or incentives (Harrison-Walker and Jiang 2023). Some businesses engage in this practice to enhance their reputation or damage competitors' reputations (Banerjee and Chua 2023). For companies, fake reviews present both risks and opportunities. They must navigate the challenge of using user-generated content for marketing while ensuring the authenticity of the information they present (Lie et al. 2024; Sahut et al. 2024).

Researchers have recognized the importance of detecting fake reviews. However, recent literature reviews on the topic (e.g., Sahut et al. 2024; Walther et al. 2023; Wu et al. 2020) highlight the complex nature of this task. Efforts to identify them involve a combination of manual inspection and automated systems that leverage machine learning and AI algorithms (Sahut et al. 2024). According to Walther et al. (2023), five key areas for detecting fake reviews include: (1) characteristics of the review, (2) textual features, (3) traits of the reviewer, (4) attributes of the seller, and (5) features of the platform. Wu et al. (2020) note that motivations behind review manipulation often include seeking rewards and fulfilling psychological needs, such as reducing frustration, showing self-declared brand loyalty, and boosting social status. However, the antecedents of fake reviews remain largely unexplored (Sahut et al. 2024). Psychological and motivational factors driving consumers to engage in this behavior deserve further investigation (Birim et al. 2022; Wu et al. 2020). In particular, personality traits may influence such behaviors (Harrison et al. 2018).

In a recent study, Kapoor et al. (2021) applied MDT (Bandura 1991, 1999) and examined Dark Triad personality traits to understand consumers' intentions to exaggerate their consumption experiences in online reviews. Through four

studies, they found some significant positive relationships between dark personality traits and the intention to exaggerate reviews—whether positive or negative—across both search and experience products. However, the findings regarding narcissism and psychopathy were inconsistent when examining their direct effects. Specifically, psychopathy showed nonsignificant effects in three out of four experimental scenarios. Kapoor et al. (2021) addressed this inconsistency by demonstrating that moral disengagement mediates the relationship between Dark Triad traits and exaggeration intentions, effectively explaining the variability in their direct effects. However, their study focused on intentions rather than actual review-posting behavior and on exaggerated reviews rather than fake ones. While exaggerated reviews may involve overstatement, they remain grounded in real consumer experiences and represent a writing style not inherently linked to fake reviews (Banerjee 2022). Moreover, intentions associated with Dark Triad traits do not always translate into actual deceptive behaviors (Harrison et al. 2018), suggesting a need to examine whether and how these traits influence the actual behavior of posting fake reviews. These findings underscore the importance of shifting the focus from intentions to enacted deception, by further researching into the behavioral manifestations of the Dark Triad in online deception.

### 2.3 | Theoretical Framework: Interpersonal Deception Theory

Numerous theories and models have been proposed to explain deceptive behavior in online settings, including self-selection theory, information processing theory, and source credibility theory (Birim et al. 2022). However, this study adopts IDT, developed by Buller and Burgoon (1996), as it offers a comprehensive framework for understanding deception in interpersonal communication. IDT defines deception as “a message knowingly transmitted by a sender to foster a false belief or conclusion by the receiver” (Buller and Burgoon 1996, 205). This occurs when communicators strategically manipulate information to create meanings that deviate from the truth. The process requires significant cognitive effort. The theory posits that deception involves both strategic and nonstrategic behaviors. Strategic behaviors are deliberate actions aimed at manipulating messages, managing impressions, and fostering false beliefs, while nonstrategic behaviors, or “leakage,” are involuntary cues that may unintentionally reveal deceptive intentions due to the cognitive load involved in sustaining the deception (Buller and Burgoon 1996). Consequently, effective interpersonal communication is viewed as a skillful activity, where understanding the dynamics between participants is crucial for accurately interpreting communication phenomena (Buller and Burgoon 1996).

The term “interpersonal” in IDT refers to the dynamic exchange of messages between two or more individuals, which is not necessarily interactive (Buller and Burgoon 1996). While traditional applications of IDT have focused on face-to-face interactions, recent studies (e.g., Birim et al. 2022; Zhang et al. 2016) suggest that its principles can be effectively extended to online contexts, where communication often occurs asynchronously and without direct interaction. For instance, deceptive online reviews may lack the nonverbal cues

characteristic of face-to-face communication but can still reveal signs of deception through specific language patterns, such as overly detailed descriptions, excessive emotional appeals, or inconsistencies in the narrative (Zhang et al. 2016).

IDT highlights deception as a goal-oriented activity, strategically crafted to manipulate the receiver's beliefs. Broadly, senders employ deception to fulfill instrumental, relational, and identity objectives, which may be shaped by their personality traits (Buller and Burgoon 1996). For example, instrumental motivations suggest crafting a fake review to maintain influence over the receiver, aligning with IDT's focus on the deliberate and purposeful nature of deceptive communication. These insights make IDT particularly relevant for studying deceptive behaviors in online consumer reviews.

This study adopts IDT to understand how individual personality traits, influence the behavior of posting fake reviews. By applying the principles of IDT to the distinct characteristics of the Dark Triad traits, we can conceptualize how goal-oriented and strategic deceptive behaviors are shaped by individual tendencies. Unlike exaggerated reviews, which are grounded in real experiences, fake reviews involve deliberate fabrication and strategic manipulation, aligning closely with IDT's conceptualization of deception. The applicability of IDT to the online domain has been demonstrated in prior research (e.g., Birim et al. 2022; Zhang et al. 2016), offering a robust foundation for analyzing deceptive practices. This study builds on these findings to focus specifically on the strategic behaviors outlined by IDT in the context of fake reviews.

## 2.4 | Machiavellianism and Fake Reviews

Machiavellians are more likely to engage in deceptive behaviors such as lying, stealing from, cheating and misleading others (Trevino and Youngblood 1990; Harrison et al. 2018; Kapoor et al. 2021). To influence others and achieve their personal goals, Machiavellians utilize a set of manipulative tactics. The latter include ingratiation, supplication, and self-disclosure, which allow them to appear more likeable and trustworthy to others (Jonason and Webster 2012). By leveraging these tactics, they can craft favorable impressions and manipulate social interactions to their advantage.

In offline conversations Machiavellians have been associated with complex forms of socially unacceptable manipulation, such as deliberately lying (Wilson et al. 1996), using entirely fabricated information to create plausible messages. In the realm of online reviews, these manipulative tendencies may manifest as fake reviews, where Machiavellians fabricate experiences that—distinctively from exaggerated reviews (Kapoor et al. 2021)—are not grounded in any actual consumer experiences, to mislead potential consumers and influence their purchasing decisions. Posting fake reviews aligns with Machiavellians' inclination to use deception for personal gain and their belief that manipulation is a valid means of achieving their goals (O'Boyle et al. 2012). Their propensity toward unethical behaviors further supports the likelihood that they would consider posting fake reviews as a means of manipulating

consumer perceptions and influencing market outcomes (Harrison et al. 2018; Drouin et al. 2016).

According to IDT, deceivers must strategically manipulate information to create believable messages and are often driven by instrumental goals (Buller and Burgoon 1996). IDT defines "strategic behaviors" as reflecting broad plans and intentions rather than specific routines or tactics (Buller and Burgoon 1996). Despite their expertise in various manipulative tactics, it is crucial to note that Machiavellians believe interpersonal manipulation is key to life success and act accordingly (Furnham et al. 2013). Given their strong tendency for manipulation, Machiavellians fit the profile of the skilled individuals described in IDT, capable of effectively manipulating their communication behavior. Therefore, in the context of online reviews, it is reasonable to hypothesize that:

**H1.** *Machiavellianism is positively associated with the behavior of posting fake reviews.*

## 2.5 | Narcissism and Fake Reviews

Narcissism is characterized by grandiosity, self-focus, and self-importance, along with a strong need for validation and attention (Morf and Rhodewalt 2001). Narcissists often exhibit entitlement, self-absorption, and a willingness to exploit others (Emmons 1987). While narcissists may initially be perceived favorably, their tendencies toward arrogance and exploitation become more apparent over time (Paulhus 1998). This behavior is goal-oriented and aimed at obtaining affirmation, often insensitive to social constraints (Morf and Rhodewalt 2001). Narcissists seek social approval and admiration, using public platforms to showcase their experiences and elevate their social status (Besser et al. 2016). They often post reviews online as a form of self-representation and to project their decision-making ability (Kapoor et al. 2021).

However, while narcissists engage in self-serving and at times unethical behaviors, they strategically avoid actions that could jeopardize their long-term social standing and credibility. As Bergman et al. (2011, 706) note, narcissists "lack empathy and have few, if any, close relationships, yet they strongly desire social contact, as others serve as their primary source of admiration and attention." Unlike Machiavellianism and psychopathy, narcissism is uniquely characterized by a strong focus on maintaining a favorable self-image and social superiority (Jones and Paulhus 2011). This leads narcissists to engage in behaviors that reinforce their perceived expertise and authenticity, rather than those that could publicly expose them as deceptive. For this reason, posting fake reviews could threaten their credibility and authenticity in the online community, which are essential to their self-image.

According to IDT, "deceivers may experience varying degrees of physiological arousal and negative affect stemming from detection apprehension (the fear of being caught deceiving) and guilt or discomfort associated with violating conversational rules and social prescriptions against deceit" (Buller and Burgoon 1996, 10). For narcissistic individuals, this detection

apprehension is particularly pronounced in contexts where deception could damage their self-presentation. Posting fake reviews—if discovered—could severely undermine their perceived credibility and expertise, which are central to their self-image. Additionally, due to their primary needs being genuine validation and self-enhancement, Jones and Paulhus (2017) also found that narcissists are less likely to engage in risky deception activities compared to other Dark Triad personality traits. Therefore, we hypothesize that:

**H2.** *Narcissism is negatively associated with the behavior of posting fake reviews.*

## 2.6 | Psychopathy and Fake Reviews

Psychopathic traits are characterized by emotional callousness, impulsivity, and inadequate self-control (Paulhus and Williams 2002). Individuals with these traits frequently engage in behaviors driven by immediate egocentric gratification, often making impulsive, short-term decisions (Harrison et al. 2018). While they may engage in deceptive behaviors, these are typically spontaneous rather than strategically planned (Anestis et al. 2009). Posting fake reviews, however, involves a deliberate and calculated form of deception. According to IDT, deception in interpersonal contexts requires careful planning and manipulation of information to maintain credibility, which is more cognitively demanding than truthful communication (Buller and Burgoon 1996). IDT differentiates between spontaneous deception and strategic deception, with the latter requiring effort, consistency, and the ability to sustain a believable narrative (Xiao and Benbasat 2011). Crafting fake reviews falls under the strategic category, as it requires constructing detailed, plausible content and maintaining credibility over time (Birim et al. 2022).

The Triarchic Model of Psychopathy (Patrick et al. 2009) outlines three core traits: boldness, meanness, and disinhibition. While meanness is associated with exploitative and harmful behaviors, boldness and disinhibition are characterized by high self-assurance and impulsivity, respectively. These traits suggest that individuals with psychopathic tendencies prioritize immediate gratification over calculated, goal-directed deception. Due to their impulsivity, they are more likely to engage in reactive, unstructured dishonesty rather than effortful and sustained deception (Jones and Paulhus 2011), which undermines their capacity for the strategic behavior IDT describes. Furthermore, the short-term focus inherent in psychopathy (Harrison et al. 2018) conflicts with the deliberate, sustained effort IDT emphasizes for successful deception (Petrescu et al. 2022). Psychopaths tend to favor spontaneous, uncalculated actions (Ramírez-Martín et al. 2020) over the calculated manipulation necessary for creating credible fake reviews (Buller and Burgoon 1996). For this reason, their impulsive nature and short-term focus make them less likely to engage in the cognitively demanding task of posting fake reviews. Hence, we hypothesize that:

**H3.** *Psychopathy is negatively associated with the behavior of posting fake reviews.*

## 3 | Methods

### 3.1 | Data Set

For this study, we utilized Amazon's fake product reviews data set, a widely recognized resource in the domain of deception detection in online reviews (Birim et al. 2022; Hajek et al. 2023; Kokkodis et al. 2022; Kronrod et al. 2023). This data set, originally sourced from Aayush Saxena's GitHub repository (Saxena 2018), comprises 21,000 reviews, evenly split between 10,500 labeled as fake and 10,500 as genuine (Birim et al. 2022). Each review is annotated with features including product ID, product category, verified purchase status, star rating, and a label indicating the authenticity of the review (Hajek et al. 2023).

The data set is particularly valuable for several reasons. First, it provides a balanced distribution of fake and genuine reviews. This balance helps ensure that algorithms can learn to distinguish between real and fake reviews effectively, without being biased toward one class (Kokkodis et al. 2022). Second, the inclusion of verified purchase data adds another dimension to the analysis by allowing for the exploration of whether purchasing behavior influences the likelihood of writing real or fake reviews (Birim et al. 2022). The reviews span 30 different product categories, with each category containing 700 reviews—350 fakes and 350 reals. This diversity in product categories enhances the data set's applicability to various domains within e-commerce, making it a robust tool for studying the patterns and characteristics of fake reviews across different contexts (Kronrod et al. 2023).

Amazon was selected for this study due to its dominant position in the online retail sector and the comprehensive nature of its review system, which includes a mix of positive and negative reviews with a full range of star ratings (Birim et al. 2022). The platform's popularity and extensive review database make it a fertile ground for both genuine consumer feedback and the potential for fraudulent activity. This environment provides a realistic setting for studying deceptive behaviors in reviews, as well as the challenges of identifying such behaviors (Hajek et al. 2023).

### 3.2 | Variables

#### 3.2.1 | Dependent Variable

The dependent variable in this study is a binary variable indicating whether a review is categorized as fake (*Fake Review*). This variable takes the value of 1 if the review is identified as fake and 0 otherwise. The categorization method aligns with established methodologies in the literature, as utilized by Birim et al. (2022), Hajek et al. (2023), and Elmurngi and Gherbi (2017). This binary categorization facilitates the clear delineation between authentic and deceptive reviews, enabling a focused analysis of the linguistic and psychological traits that characterize fake reviews.

### 3.2.2 | Focal Independent Variables

To investigate the influence of Dark Triad personality traits on framing deception behaviors in writing fake reviews, we utilize linguistic characteristics identified through the Linguistic Inquiry and Word Count (LIWC) tool (Pennebaker et al. 2015). Our approach is informed by the methodology outlined by Yousaf and Kim (2023), which provides a robust empirical foundation for the linguistic markers associated with each Dark Triad trait, supported by extensive prior research.

#### Machiavellianism:

Nazari et al. (2015) and Sumner et al. (2012) found that Machiavellianism is associated with a higher use of swear words and negative emotional words, as well as a lower use of positive emotional words. Additionally, Muir et al. (2016) demonstrated that Machiavellians may adjust their linguistic style based on social dynamics, leading to shorter texts (Bogolyubova et al. 2018). These linguistic tendencies align with Machiavellians' manipulative nature, as their language often conveys cynicism, distrust, and strategic negativity aimed at influencing others (Sumner et al. 2012). These linguistic traits are combined to create our Machiavellianism index, using the following formula:

$$\text{Machiavellianism} = \text{Swear Words} + \text{Negative Emotions} + \frac{1}{1 + \text{Positive Emotions}} + \frac{1}{1 + \log(\text{WordCount})}$$

#### Narcissism:

Holtzman et al. (2019), building on Carey et al. (2015), identified linguistic patterns such as the use of second-person pronouns and swear words, as well as a tendency to avoid anxiety-related, tentative, and perceptual process-related words. While narcissism is traditionally linked to self-focus, linguistic studies (Holtzman et al. 2019) suggest that narcissists often use second-person pronouns to engage their audience directly (Bergman et al. 2011), seeking admiration and reinforcing their dominance in social interactions. Moreover, Bogolyubova et al. (2018) found that narcissists tend to write longer texts. The narcissism index is calculated as follows:

$$\text{Narcissism} = \text{2nd Person Pronouns} + \text{Swear Word} + \frac{1}{s + \log(\text{WordCount}) + 1 + \text{Anxiety}} + \frac{1}{1 + \text{Tentative}} + \frac{1}{1 + \text{Perceptual Processes}}$$

#### Psychopathy:

Hancock et al. (2018) demonstrated that individuals with high levels of psychopathy are characterized by a focus on the past and an increased use of discrepancy-related, anger, and swear words. The authors found that individuals with psychopathic tendencies demonstrate a high focus on past events with the

purpose of reflecting a detached, narrative style when recalling past events for their (re)actions. This is consistent with research showing that psychopaths tend to justify or rationalize past behaviors rather than express concern for future consequences due to their short-term focus nature (Harrison et al. 2018). They tend to avoid using present tense and first-person singular pronouns, contributing to the psychopathy index calculated as follows:

$$\text{Psychopathy} = \text{Past Focus} + \text{Swear Words} + \text{Discrepancy}$$

$$+ \text{Anger} + \frac{1}{1 + \text{1st Person Singular Pronouns}} + \frac{1}{1 + \text{Present Focus}}$$

The three main independent variables are nonnegative continuous variables. By leveraging these formulas devised by Yousaf and Kim (2023), as well as foundational studies (e.g., Bogolyubova et al. 2018; Hancock et al. 2018; Holtzman et al. 2019; Nazari et al. 2015; Sumner et al. 2012), we operationalize the Dark Triad traits through specific linguistic markers in the text of reviews. It is important to note that the indices for Machiavellianism, Narcissism, and Psychopathy are derived using trait-specific formulas based on linguistic features. These indices serve as relative measures of the presence of linguistic patterns associated with each trait in the data set. Because the formulas for each trait are composed of different variables and scales, the scores are not directly comparable across traits. Instead, they should be interpreted as reflecting the relative levels of each trait within the context of the data set.

### 3.2.3 | Control Variables

Consistent with the existing literature examining fake reviews, a variety of control variables are incorporated into the model to account for potential confounding factors that could influence the relationship between the Dark Triad traits and the likelihood of a review being fake. These control variables ensure a more robust and accurate analysis by controlling for factors related to the review content, reviewer behavior, and product characteristics.

To control for the influence of review length, we include the number of words in each review (*WordCount*). Previous studies (e.g., Birim et al. 2022; Hajek et al. 2023; Elmurngi and Gherbi 2017) have indicated that the length of a review can affect its perceived credibility and informativeness. By accounting for the number of words, we mitigate any bias related to the verbosity of the reviewer. The rating given by the reviewer to the product is another control variable (*Review Rating*), as it can affect the likelihood of a review being perceived as fake (Elmurngi and Gherbi 2017). Reviews with extreme ratings might be more likely to be questioned for authenticity. Furthermore, we control for the sentiment of the review text (*Sentiment Score*), which allows us to isolate the effect of opinions and private states (such as feelings, speculations and beliefs) of the review from other linguistic markers of deception. In line with Birim et al. (2022), we perform sentiment analysis using the VADER tool which provides a sentiment score for each review, capturing its overall sentiment

polarity. VADER, a lexicon-based sentiment classifier recognized for its high accuracy in various domains including social media (Hutto and Gilbert 2014), assigns a polarity score indicating how positive, negative, or neutral a review is.

In addition to sentiment, we control for the emotional tone of the review (*Emotional Tone*), measured using the LIWC tool (Moon et al. 2021). Emotional tone captures the overall emotional style of the review, with higher values indicating a positive, upbeat style and lower values suggesting greater anxiety, sadness, or hostility (Pennebaker et al. 2015). This measure is distinct from the sentiment score as it reflects the general emotional expressiveness of the review, allowing us to account for variations in emotional intensity. We also control for whether the review is from a verified purchase on Amazon (*Verified Purchase*). Verified purchase reviews are generally more credible, potentially affecting their categorization as fake or genuine (Hajek et al. 2023). Analytical thinking is controlled using scores from the LIWC tool (*Analytical Thinking*), following the approach of Alsubari et al. (2020). This score reflects the cognitive complexity of the review, with higher values indicating more formal, logical, and hierarchical thinking, while lower values suggest more informal and narrative thinking (Pennebaker et al. 2015). This control helps isolate the impact of cognitive style on the review's authenticity. In this regard, both the Emotional Tone and Analytical Thinking summary variables have been scaled to a 100-point scale by LIWC, ranging from 0 to 100, to improve their interpretability (Pennebaker et al. 2015).

Finally, we control for product category by creating a series of dummy variables representing the 30 product types included in our data set (Hajek et al. 2023). This accounts for potential variations in review characteristics across different product categories, ensuring that the influence of Dark Triad traits on deceptive behavior is not confounded by the type of product being reviewed.

### 3.3 | Empirical Model

To empirically test our hypotheses, we employed a logistic regression model, appropriate given the binary nature of our dependent variable: whether a review is fake or not (Wooldridge 2010). The logistic model is particularly suited for this type of binary outcome as it estimates the probability of an event occurring, in this case, the likelihood of a review being fake.

The structure of our econometric model is specified as follows:

$$\begin{aligned}
 \text{Fake Review} = & \beta_0 + \beta_1 \text{Machiavellianism} + \beta_2 \text{Narcissism} \\
 & + \beta_3 \text{Psychopathy} \\
 & + \beta_4 \text{Log(WordCount)} + \beta_5 \text{Sentiment Score} + \beta_6 \\
 & \text{Analytical Thinking} + \beta_7 \text{Emotional Tone} \\
 & + \beta_8 \text{Review Rating} + \beta_9 \text{Verified Purchase} \\
 & + \theta' \text{Product Category} + \varepsilon
 \end{aligned}$$

where  $\varepsilon$  represents the error term, capturing all unobserved factors that may affect the probability of a review being fake. In this model, the coefficients  $\beta_i$  represent the change in the log-odds of a review being fake associated with a one-unit change in the corresponding independent variable, holding all other variables constant. This model allows us to estimate the influence of the Dark Triad traits, along with other covariates, on the likelihood of posting a fake review. The descriptive sample statistics are included in Table 2. The average scores for Machiavellianism, Narcissism, and Psychopathy, derived from linguistic markers using the formulas outlined in our methodology, represent relative measures that are not directly comparable across traits, as noted in Section 3. That said, reviews in our sample exhibit a relatively high rating score (mean = 4.128) and a generally positive and upbeat tone, reflected in the high average Emotional Tone score (mean = 74.438). Analytical Thinking scores (mean = 52.397) highlight a balanced complexity of cognition among reviewers in our data set. Additionally, a comprehensive correlation table is presented in Table 3, illustrating the relationships among the dependent variable, independent variables (including the Dark Triad traits), and control variables. Upon examining the correlation values, none of the coefficients exceeds 0.7 (absolute value), based on Ratner (2009) threshold, indicating that multicollinearity is not a concern in this data set.

## 4 | Findings

Table 4 displays the empirical results of our analysis. To address potential multicollinearity concerns, we assessed the variance inflation factor values. These ranged from 1.02 to 2.79, remaining comfortably below the commonly accepted threshold of 10 (Hair et al. 1992), indicating that multicollinearity is not a significant issue in our estimates.

Model 0 serves as the baseline model, incorporating only control variables. This initial model provides a foundation for understanding the influence of Machiavellianism, narcissism, and psychopathy on fake review posting behaviors. Building on this, Model 1 examines the effect of Machiavellianism. Consistent with Hypothesis 1, we find a positive and significant relationship ( $\beta = 0.0342, p < 0.01$ ). Thus, individuals with higher Machiavellian tendencies are more likely to engage in this behavior.

Model 2 introduces narcissism as an independent variable. Supporting Hypothesis 2, the analysis reveals a negative and significant association between narcissism and fake review posting ( $\beta = -0.0192, p < 0.05$ ), suggesting that narcissistic individuals are less inclined to participate in this activity.

Model 3 investigates the role of psychopathy. The coefficient related to the association between psychopathy and fake reviews shows a negative and significant value ( $\beta = -0.0241, p < 0.001$ ), in line with Hypothesis 3.

Finally, following Furnham et al.'s (2013) prescription, Model 4 provides a comprehensive perspective by examining the simultaneous influence of all three Dark Triad

**TABLE 2** | Descriptive sample statistics.

	Mean	SD	Min	Max
Fake Review	0.500 <sup>a</sup>	0.500	0	1
Machiavellianism	1.672	2.038	0.239	25.057
Narcissism	6.809	2.002	3.250	26.166
Psychopathy	6.183	4.416	0.091	35.330
Log (WordCount)	3.931	0.707	1.609	7.974
Sentiment Score	0.580	0.505	-1.000	0.999
Analytical Thinking	52.397	28.758	1.000	99.000
Emotional Tone	74.438	32.317	1.000	99.000
Review Rating	4.128	1.278	1	5
Verified Purchase	0.557 <sup>a</sup>	0.497	0	1
Observations		21,000		

<sup>a</sup>in percentage terms.

personality traits—Machiavellianism ( $\beta = 0.0448, p < 0.001$ ), Narcissism ( $\beta = -0.0235, p < 0.01$ ), and Psychopathy ( $\beta = -0.0262, p < 0.001$ )—on fake review posting behavior. This simultaneous analysis allows us to consider the unique contribution of each trait in the presence of the others. Notably, Model 4 has the lowest AIC value among the models tested, indicating its optimal fit for the data. The results align with those of the individual models, reinforcing the distinct contributions of each trait and confirming the robustness of our findings (Miller et al. 2019).

In addition to these personality traits, our models control for several key factors influencing fake review behavior. For instance, the length of the review shows a significant negative relationship across all models ( $p < 0.001$ ), suggesting that more detailed reviews are less likely to be fake (Birim et al. 2022). Sentiment scores and emotional tone also play significant roles, indicating that both the content and tone of reviews are critical factors in detecting potential deceit (Hajek et al. 2023; Moon et al. 2021). Moreover, “Verified Purchase” labels are strongly associated with a lower likelihood of fake reviews ( $p < 0.001$ ), reinforcing the credibility these labels confer in the eyes of consumers (Hajek et al. 2023).

To ensure the robustness of our findings, we conducted additional analyses excluding log (WordCount) and Emotional Tone as control variables, given their potential overlap with the linguistic markers used to measure Dark Triad traits. The results of these alternative specifications confirm that the effects of Machiavellianism, narcissism, and psychopathy remain statistically significant and directionally stable, reinforcing the validity of our main model. Furthermore, Model 4 was selected as the final model based on its superior fit, as indicated by the lowest AIC value among all tested models. Additionally, as a further robustness check, we tested all the models using the probit estimation technique, the results of which are included in Appendix S1. The probit model results corroborate the findings from the logit models, affirming the reliability and validity of our conclusions.

Overall, these findings provide substantial support for our three hypotheses, highlighting the distinct influences of the Dark

Triad traits on fake review posting behavior. The consistent significance levels and directional effects across models affirm the robustness of our results, contributing valuable insights into the personality antecedents of posting fake reviews.

## 5 | Discussion, Implications, and Further Research

To date, research has highlighted the impact of Dark Triad traits—Machiavellianism, narcissism, and psychopathy—on various unethical behaviors (Harrison et al. 2018; Jones 2014; Jones and Paulhus 2017; Miller et al. 2019). However, the distinct characteristics of these individual traits that may differently drive the actual behavior of posting fake reviews remained underexplored. This study investigates how Dark Triad traits independently and concurrently influence this behavior. Grounded in IDT and informed by literature on the Dark Triad, we hypothesise unique effects for Machiavellianism, narcissism, and psychopathy. The findings confirm our hypotheses, revealing the significant effects of these traits on the behavior of posting fake reviews. The next section delves into these new insights and their theoretical implications.

### 5.1 | Theoretical Implications

Our study makes several theoretical contributions to the literature at the intersection of psychology, consumer behavior, and eWOM. First, while previous research (Kapoor et al. 2021) utilized MDT (Bandura 1991, 1999) to explain how Dark Triad traits influence consumers' intentions to exaggerate in online reviews, our study is the first to apply IDT (Buller and Burgoon 1996) to examine the effects of these traits on reviewers' actual behavior of posting fake online reviews. As highlighted by Birim et al. (2022), IDT focuses on deception in interpersonal communication, which directly aligns with the behavior of posting fake reviews. By shifting from MDT's focus on cognitive engagement and moral justification to IDT's emphasis on strategic deceptive behaviors, our study extends prior

TABLE 3 | Correlation table.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
(1)	Fake Review	1								
(2)	Machiavellianism	-0.00122	1							
(3)	Narcissism	-0.0221**	0.0800***	1						
(4)	Psychopathy	-0.0447***	0.235***	-0.00955	1					
(5)	Log (WordCount)	-0.0760***	0.00341	0.308***	-0.0294***	1				
(6)	Sentiment Score	0.0198**	-0.495***	0.0156*	-0.200***	0.111***	1			
(7)	Analytical Thinking	-0.151***	-0.0169*	-0.00996	-0.144***	0.147***	0.0143*	1		
(8)	Emotional Tone	0.0489***	-0.647***	-0.0680***	-0.214***	-0.137***	0.686***	-0.0430***	1	
(9)	Review Rating	-0.00980	-0.308***	-0.00205	-0.201***	-0.0226***	0.534***	0.0220***	0.447***	1
(10)	Verified Purchase	-0.570***	-0.0326***	-0.0483***	0.0261***	-0.105***	0.0211***	0.0547***	0.0272***	0.102***

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

research by providing a more behaviorally grounded understanding of fake review posting. This approach contributes novel insights into how Dark Triad traits manifest in enacted deception, offering a complementary perspective to prior studies that have primarily examined deceptive intentions.

Second, through the lens of IDT, our study provides new insights into the distinct effects of each Dark Triad trait, as recommended by Miller et al. (2019). In doing so, we also respond to the call for more research on the antecedents of fake reviews (Sahut et al. 2024), particularly those linked to psychological motives (Birim et al. 2022; Wu et al. 2020). This study extends IDT and the notion of strategic behaviors—rooted in the goal-oriented nature of interpersonal communication—to the context of online communication, specifically the act of posting fake reviews. Adding to our existing knowledge about various manipulative tactics of Machiavellian individuals (e.g., Furnham et al. 2013; Harrison et al. 2018; Kapoor et al. 2021), our findings suggest that writing fake online reviews, intended to shape public perceptions, is a manipulative strategy commonly employed by Machiavellians in digital contexts. Consistent with the instrumental motives that drive intentional acts of deception (Buller and Burgoon 1996), Machiavellians tend to consider posting fake reviews as a deliberate means of maintaining influence over other consumers' perceptions.

Moreover, the findings of a significant negative relationship between narcissism among online reviewers and their propensity to post fake reviews suggest that detection apprehension—the concern about being caught (Buller and Burgoon 1996)—motivates narcissists to control deceptive behavior online. This is driven by their primary needs for genuine validation and self-enhancement (Jones and Paulhus 2017). Although these findings contrast with Kapoor et al. (2021), who identified narcissism as a predictor of the intention to write exaggerated online reviews, they complement the idea that deception serves identity-related motivations, benefiting the principal beneficiary of deceit (Buller and Burgoon 1996). Specifically, the desire for external admiration and social approval (Besser et al. 2016) may lead narcissists to view online reviews as a form of self-presentation (Kapoor et al. 2021). However, the potential consequences of being caught—such as platform bans—pose a significant threat to their online presence, resulting in a loss of attention and admiration from others. Therefore, our findings demonstrate that the risk of losing credibility and authenticity within online communities, which are crucial to their self-image, effectively deters narcissists from posting fake reviews.

Challenging prior assumptions that psychopaths broadly engage in unethical behavior (e.g., Harrison et al. 2018; Kapoor et al. 2021), our study provides empirical evidence suggesting that the impulsive and short-term focus characteristic of psychopathy does not align with the calculated and strategic effort required to post fake reviews. Supporting the idea that deception is a cognitively demanding task—requiring more effort to craft plausible messages than to create truthful ones (Anestis et al. 2009; Patrick et al. 2009)—we reveal a negative association between psychopathy and fake review posting. This finding offers new insights into the previously inconclusive results on

**TABLE 4** | Estimation results of logit models.

	<b>Model 0</b>	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
Machiavellianism		0.0342** (0.0108)			0.0448*** (0.0111)
Narcissism			-0.0192* (0.00892)		-0.0235** (0.00897)
Psychopathy				-0.0241*** (0.00387)	-0.0262*** (0.00390)
Log (WordCount)	-0.520*** (0.0288)	-0.513*** (0.0289)	-0.503*** (0.0298)	-0.525*** (0.0289)	-0.497*** (0.0300)
Sentiment Score	0.119* (0.0517)	0.128* (0.0518)	0.119* (0.0517)	0.114* (0.0518)	0.125* (0.0518)
Analytical Thinking	-0.0104*** (0.000609)	-0.0104*** (0.000610)	-0.0105*** (0.000611)	-0.0110*** (0.000620)	-0.0111*** (0.000622)
Emotional Tone	0.00216** (0.000747)	0.00349*** (0.000869)	0.00213** (0.000747)	0.00166* (0.000753)	0.00332*** (0.000872)
Review Rating	0.0794*** (0.0163)	0.0794*** (0.0163)	0.0800*** (0.0163)	0.0693*** (0.0164)	0.0692*** (0.0164)
Verified Purchase	-2.992*** (0.0401)	-2.992*** (0.0401)	-2.994*** (0.0401)	-2.987*** (0.0401)	-2.988*** (0.0402)
Product Category Dummies	Included	Included	Included	Included	Included
Constant	4.000*** (0.170)	3.811*** (0.181)	4.070*** (0.173)	4.278*** (0.176)	4.139*** (0.187)
Observations	21,000	21,000	21,000	21,000	21,000
Pseudo $R^2$	0.292	0.292	0.292	0.293	0.294
AIC	20,679.2	20,672.0	20,676.7	20,646.2	20,629.8
Log Likelihood	-10,303.6	-10,299.0	-10,301.4	-10,286.1	-10,275.9

Note: Standard errors are in parentheses.

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

<sup>†</sup> $p < 0.1$ .

psychopathy's role in the online review context (e.g., Kapoor et al. 2021). Specifically, our study demonstrates that psychopathic individuals do not perceive others as relevant enough to influence their behavior. As a result, their online deception is unlikely to be motivated by relational goals, such as conforming to the expectations of online communities. Instead, they are less likely to pay attention to how the genuine reviews of others were crafted. In combination with impulsivity, short-term focus, and reduced ability to articulate thoughts (Patrick et al. 2009; Bogolyubova et al. 2018), they are less likely to take time to recognize and adhere to social norms that could help successfully craft plausible fake reviews. This explanation aligns, to some extent, with findings by Yousaf and Kim (2023), who found a negative impact of psychopathy on review helpfulness.

Lastly, building on the eWOM literature addressing online content authenticity (e.g., Lie et al. 2024; Zhang et al. 2016), our study expands the use of NLP techniques (e.g., Birim et al. 2022; Hajek et al. 2023) to identify the linguistic markers of Dark Triad traits in fake reviews on digital retail platforms like

Amazon. Extending the work of Yousaf and Kim (2023), who linked the Dark Triad to review ratings and perceived helpfulness, our findings reveal its direct connection to fake review posting, shedding light on a critical yet underexplored aspect of online deception. While prior research has shown that review authenticity positively impacts the helpfulness perceived by consumers (Lie et al. 2024) and that Machiavellian reviewers drive consumers' perceptions of review helpfulness (Yousaf and Kim 2023), our study offers a new implication. Our findings suggest that Machiavellian reviewers may also craft fake reviews that appear both authentic and helpful. This creates a paradox where the same traits that foster perceptions of review helpfulness can be used to deceive consumers, undermining trust in eWOM systems. This underscores the critical need for advanced detection mechanisms to safeguard the integrity of digital marketplaces. By examining personality factors influencing review authenticity—alongside valence and helpfulness, which are key drivers of consumer decisions (Lie et al. 2024)—our study significantly enhances understanding of deceptive practices in eWOM and their implications for trust in digital marketplaces.

## 5.2 | Managerial Implications

This study provides several practical insights for managers and platform developers, particularly those responsible for platforms or businesses heavily dependent on online reviews. First, our findings can enhance the effectiveness of fake review detection systems. Specifically, online review platform managers should incorporate metrics that identify Dark Triad personality traits within their detection models. As our results demonstrate, this approach could significantly improve the explanatory power of these models, enabling businesses to identify fake reviews with greater accuracy. While textual analysis, especially sentiment analysis (e.g., Birim et al. 2022; Hajek et al. 2023), has proven essential in prior research, we advocate for extending text analytics to also capture personality traits. Businesses should invest in advanced technologies capable of filtering out fake reviews, thereby promoting authentic feedback.

Second, recognizing that individuals with Machiavellianism are more likely to post fake reviews, platform managers should tailor their strategies to identify and mitigate this behavior. This could involve developing detailed user profiles based on linguistic patterns indicative of Machiavellianism. Reviews flagged for high Machiavellianism should undergo additional scrutiny, focusing on users with multiple reviews exhibiting these traits. A deeper investigation into such profiles could help distinguish between strategic and nonstrategic deceptive behaviors, allowing for a more comprehensive approach to combatting fake reviews.

Third, the study reveals that individuals with narcissism are less inclined to post fake reviews, likely due to their desire for authentic recognition and concern over damaging their social standing. Businesses could leverage this insight by promoting a culture of genuine feedback, emphasizing how credible reviews enhance the reviewer's reputation as a reliable source. However, it remains crucial to maintain a high level of deterrence against fake reviews by reinforcing the consequences of such behavior (e.g., banning the reviewer from the platform or imposing fines).

Fourth, our findings indicate that the higher the review sentiment and review rating, the greater the likelihood of the review being fake. This suggests that platform managers should be cautious when interpreting high ratings and overly positive sentiments, as they may not always reflect authentic customer experiences. Monitoring sentiment trends (Hajek et al. 2023) and applying skepticism to high-scoring reviews could enhance the credibility of platforms and improve decision-making for potential customers. Moreover, consistent with prior research (e.g., Birim et al. 2022), our findings highlight the importance of verified purchase reviews as a critical cue for identifying fake reviews. Platforms should prioritize making this information readily accessible to consumers to enhance trust and decision-making.

Overall, by understanding the specific personality traits that influence online behaviors, businesses can better tailor their strategies for managing reviews and customer feedback. This approach will help maintain a credible online presence and

foster more authentic and trustworthy relationships with customers.

## 5.3 | Limitations and Future Research

This study is not without its limitations. Although we utilized a widely recognized data set in the field of deception detection in online reviews (Birim et al. 2022; Hajek et al. 2023; Kokkodis et al. 2022; Kronrod et al. 2023), our data lack background information about the reviewers, such as demographic details and time stamps. These constraints limit our ability to fully capture the dynamic aspects of online behavior and contextual factors that may influence the posting of fake reviews. While the formulas devised by Yousaf and Kim (2023) leverage empirically tested metrics from previous studies (e.g., Bogolyubova et al. 2018; Hancock et al. 2018; Holtzman et al. 2019; Nazari et al. 2015; Sumner et al. 2012), further validation of our NLP procedures would strengthen their methodological rigor. Future research could enhance confidence in the generalizability and accuracy of these NLP-based measures by conducting empirical studies that test linguistic markers alongside objective measures of Dark Triad traits, such as third-party assessments (Cragun et al. 2020).

Future research could enhance our understanding of the Dark Triad's impact on online behaviors. For example, longitudinal studies could further validate our findings and examine whether the influence of Dark Triad traits on online behavior remains consistent or varies over time. Such studies would provide valuable insights into the temporal stability of these traits' effects on review behavior. Additionally, future research could explore whether other reviewer characteristics (e.g., gender and age) moderate the relationship between Dark Triad traits and fake review behavior, offering a more comprehensive view of the mechanisms driving deceptive behaviors associated with the Dark Triad. Moreover, while our study identifies significant associations between Dark Triad traits and fake review posting behaviors, it does not empirically investigate the psychological mechanisms underlying these associations. Future research could address this gap by conducting in-depth psychological or experimental studies, particularly in interactive online contexts such as social media platforms or discussion forums. These settings could provide valuable insights into the mechanisms driving these behaviors and how the differences we observed are shaped by each Dark Triad trait.

Furthermore, our study focused solely on product reviews from Amazon, limiting our understanding of how fake reviews may differ across various contexts. Future research should consider investigating the differences in fake review dynamics between experiential and material goods and services. Understanding these distinctions could provide critical insights into how fake reviews are constructed and perceived, thereby informing more tailored approaches to combatting fraudulent reviews across different sectors.

Finally, based on our study results, more research is needed to develop and test interventions aimed at reducing the prevalence of fake reviews, particularly those targeting individuals with high levels of Machiavellianism. In this way, future studies

could contribute to more effective approaches in combating unethical online behaviors.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## References

Alsubari, S. N., M. B. Shelke, and S. N. Deshmukh. 2020. "Fake Reviews Identification Based on Deep Computational Linguistic." *International Journal of Advanced Science and Technology* 29, no. 8s: 3846–3856.

Anestis, M. D., J. C. Anestis, and T. E. Joiner. 2009. "Affective Considerations in Antisocial Behavior: An Examination of Negative Urgency in Primary and Secondary Psychopathy." *Personality and Individual Differences* 47, no. 6: 668–670.

Back, M. D. 2018. "The Narcissistic Admiration and Rivalry Concept." In *Handbook of Trait Narcissism: Key Advances, Research Methods, and Controversies*, edited by A. Hermann, A. Brunell, J. Foster, 57–67. Springer.

Bandura, A. 1991. "Social Cognitive Theory of Self-Regulation." *Organizational Behavior and Human Decision Processes* 50, no. 2: 248–287.

Bandura, A. 1999. "Social Cognitive Theory: An Agentic Perspective." *Asian Journal of Social Psychology* 2, no. 1: 21–41.

Banerjee, S. 2022. "Exaggeration in Fake vs. Authentic Online Reviews for Luxury and Budget Hotels." *International Journal of Information Management* 62: 102416.

Banerjee, S., and A. Y. K. Chua. 2023. "Understanding Online Fake Review Production Strategies." *Journal of Business Research* 156: 113534.

Bergman, S. M., M. E. Fearrington, S. W. Davenport, and J. Z. Bergman. 2011. "Millennials, Narcissism, and Social Networking: What Narcissists Do on Social Networking Sites and Why." *Personality and Individual Differences* 50, no. 5: 706–711.

Besser, A., V. Zeigler-Hill, M. Weinberg, and A. L. Pincus. 2016. "Do Great Expectations Lead to Great Disappointments? Pathological Narcissism and the Evaluation of Vacation Experiences." *Personality and Individual Differences* 89: 75–79.

Birim, S. Ö., I. Kazancoglu, S. Kumar Mangla, A. Kahraman, S. Kumar, and Y. Kazancoglu. 2022. "Detecting Fake Reviews Through Topic Modelling." *Journal of Business Research* 149: 884–900.

Bogolyubova, O., P. Panicheva, R. Tikhonov, V. Ivanov, and Y. Ledovaya. 2018. "Dark Personalities on Facebook: Harmful Online Behaviors and Language." *Computers in Human Behavior* 78: 151–159.

Bowen, K. T., G. Musarra, and Y. C. Ou. 2022. "How and When Narcissism and Faith in Humanity Drive Sustainable Consumption." *Psychology & Marketing* 39, no. 9: 1706–1724.

Buller, D. B., and J. K. Burgoon. 1996. "Interpersonal Deception Theory." *Communication Theory* 6, no. 3: 203–242.

Carey, A. L., M. S. Brucks, A. C. P. Küfner, et al. 2015. "Narcissism and the Use of Personal Pronouns Revisited." *Journal of Personality and Social Psychology* 109, no. 3: e1.

Chatterjee, S., R. Chaudhuri, A. Kumar, C. Lu Wang, and S. Gupta. 2023. "Impacts of Consumer Cognitive Process to Ascertain Online Fake Review: A Cognitive Dissonance Theory Approach." *Journal of Business Research* 154: 113370.

Cragun, O. R., K. J. Olsen, and P. M. Wright. 2020. "Making CEO Narcissism Research Great: A Review and Meta-Analysis of CEO Narcissism." *Journal of Management* 46, no. 6: 908–936.

Drouin, M., D. Miller, S. M. J. Wehle, and E. Hernandez. 2016. "Why Do People Lie Online? "Because Everyone Lies on the Internet"." *Computers in Human Behavior* 64: 134–142.

Elmurngi, E., and A. Gherbi. 2017. "Detecting Fake Reviews Through Sentiment Analysis Using Machine Learning Techniques." In *In IARIA 2017 Conference on Data Analytics*, 65–72. IARIA XPS Press.

Emmons, R. A. 1987. "Narcissism: Theory and Measurement." *Journal of Personality and Social Psychology* 52, no. 1: 11–17.

Fawler, G. A. 2023. "Those 10,000 5-Star Reviews Are Fake. Now They'll Also be Illegal." *Washington Post*, June 30, 2023. <https://www.washingtonpost.com/technology/2023/06/30/fake-reviews-online-ftc/>.

Furnham, A., S. C. Richards, and D. L. Paulhus. 2013. "The Dark Triad of Personality: A 10 Year Review." *Social and Personality Psychology Compass* 7, no. 3: 199–216.

Hair, J. F., W. C. Black, B. J. Babin, and R. E. Anderson. 1992. *Multivariate Data Analysis*. New York, NY: Prentice Hall.

Hajek, P., L. Hikkerova, and J. M. Sahut. 2023. "Fake Review Detection in e-Commerce Platforms Using Aspect-Based Sentiment Analysis." *Journal of Business Research* 167: 114143.

Hancock, J. T., M. Woodworth, and R. Boochever. 2018. "Psychopaths Online: The Linguistic Traces of Psychopathy in Email, Text Messaging, and Facebook." *Media and Communication* 6, no. 3: 83–92.

Harrison, A., J. Summers, and B. Mennecke. 2018. "The Effects of the Dark Triad on Unethical Behavior." *Journal of Business Ethics* 153: 53–77.

Harrison-Walker, L. J., and Y. Jiang. 2023. "Suspicion of Online Product Reviews as Fake: Cues and Consequences." *Journal of Business Research* 160: 113780.

Hollebeek, L. D., D. E. Sprott, S. Urbonavicius, et al. 2022. "Beyond the Big Five: The Effect of Machiavellian, Narcissistic, and Psychopathic Personality Traits on Stakeholder Engagement." *Psychology & Marketing* 39, no. 6: 1230–1243.

Holtzman, N. S., A. M. Tackman, A. L. Carey, et al. 2019. "Linguistic Markers of Grandiose Narcissism: A LIWC Analysis of 15 Samples." *Journal of Language and Social Psychology* 38, no. 5–6: 773–786.

Hutto, C. J., and E. Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model For." In *Eighth International AAAI Conference on Weblogs and Social Media*, 18. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109>.

Jonason, P. K., and G. D. Webster. 2012. "A Protean Approach to Social Influence: Dark Triad Personalities and Social Influence Tactics." *Personality and Individual Differences* 52, no. 4: 521–526.

Jones, D. N. 2014. "Risk in the Face of Retribution: Psychopathic Individuals Persist in Financial Misbehavior Among the Dark Triad." *Personality and Individual Differences* 67: 109–113.

Jones, D. N., and D. L. Paulhus. 2011. "The Role of Impulsivity in the Dark Triad of Personality." *Personality and Individual Differences* 51, no. 5: 679–682.

Jones, D. N., and D. L. Paulhus. 2017. "Duplicity Among the Dark Triad: Three Faces of Deceit." *Journal of Personality and Social Psychology* 113, no. 2: 329–342.

Kapoor, P. S., B. M. S. M. Maity, and N. K. Jain. 2021. "Why Consumers Exaggerate in Online Reviews? Moral Disengagement and Dark Personality Traits." *Journal of Retailing and Consumer Services* 60: 102496.

Karampournioti, E., N. Hennigs, and K. P. Wiedmann. 2018. "When Pain is Pleasure: Identifying Consumer Psychopaths." *Psychology & Marketing* 35, no. 4: 268–282.

Kokkodis, M., T. Lappas, and G. C. Kane. 2022. "Optional Purchase Verification in e-Commerce Platforms: More Representative Product Ratings and Higher Quality Reviews." *Production and Operations Management* 31, no. 7: 2943–2961.

Kowalski, C. M., P. A. Vernon, and J. A. Schermer. 2016. "The General Factor of Personality: The Relationship Between the Big One and the Dark Triad." *Personality and Individual Differences* 88: 256–260.

Kronrod, A., I. Gordeliy, and J. K. Lee. 2023. "Been There, Done That: How Episodic and Semantic Memory Affects the Language of Authentic and Fictitious Reviews." *Journal of Consumer Research* 50, no. 2: 405–425.

Lambert, A., and J. Desmond. 2013. "Loyal Now, But Not Forever! A Study of Narcissism and Male Consumer-Brand Relationships." *Psychology & Marketing* 30, no. 8: 690–706.

Lee, K., M. C. Ashton, J. Wiltshire, J. S. Bourdage, B. A. Visser, and A. Gallucci. 2013. "Sex, Power, and Money: Prediction From the Dark Triad and Honesty-Humility." *European Journal of Personality* 27, no. 2: 169–184.

Lie, D. S., B. Sung, M. Stankovic, and F. Septianto. 2024. "How Profanity in Influences Perceived Authenticity and Perceived Helpfulness of Online Reviews: The Moderating Role of Review Subjectivity." *Decision Support Systems* 178: 114144.

Malär, L., and A. Giuffredi-Kähr. 2024. "The Dark Triad of Brand Personality: Scale Development and Validation." *Psychology & Marketing* 41, no. 11: 2728–2740.

Miller, J. D., C. Vize, M. L. Crowe, and D. R. Lynam. 2019. "A Critical Appraisal of the Dark-Triad Literature and Suggestions for Moving Forward." *Current Directions in Psychological Science* 28, no. 4: 353–360.

Moon, S., M. Y. Kim, and D. Iacobucci. 2021. "Content Analysis of Fake Consumer Reviews by Survey-Based Text Categorization." *International Journal of Research in Marketing* 38, no. 2: 343–364.

Morf, C. C., and F. Rhodewalt. 2001. "Unraveling the Paradoxes of Narcissism: A Dynamic Self-Regulatory Processing Model." *Psychological Inquiry* 12, no. 4: 177–196.

Muir, K., A. Joinson, R. Cotterill, and N. Dewdney. 2016. "Characterizing the Linguistic Chameleon: Personal and Social Correlates of Linguistic Style Accommodation." *Human Communication Research* 42, no. 3: 462–484.

Nazari, Z., G. Lucas, and J. Gratch. 2015. "Multimodal Approach for Automatic Recognition of Machiavellianism." In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 215–221. IEEE.

O'Boyle, E. H., D. R. Forsyth, G. C. Banks, and M. A. McDaniel. 2012. "A Meta-Analysis of the Dark Triad and Work Behavior: A Social Exchange Perspective." *Journal of Applied Psychology* 97, no. 3: 557–579.

Patrick, C. J., D. C. Fowles, and R. F. Krueger. 2009. "Triarchic Conceptualization of Psychopathy: Developmental Origins of Disinhibition, Boldness, and Meanness." *Development and Psychopathology* 21, no. 3: 913–938.

Paulhus, D. L. 1998. "Interpersonal and Intrapsychic Adaptiveness of Trait Self-Enhancement: A Mixed Blessing?" *Journal of Personality and Social Psychology* 74, no. 5: 1197–1208.

Paulhus, D. L., and K. M. Williams. 2002. "The Dark Triad of Personality: Narcissism, Machiavellianism, and Psychopathy." *Journal of Research in Personality* 36, no. 6: 556–563.

Pennebaker, J. W., R. L. Boyd, K. Jordan, and K. Blackburn. 2015. *The Development and Psychometric Properties of LIWC2015*. Austin, TX: Pennebaker Conglomerates. [www.LIWC.net](http://www.LIWC.net).

Petrescu, M., P. Kitchen, C. Dobre, et al. 2022. "Innocent Until Proven Guilty: Suspicion of Deception in Online Reviews." *European Journal of Marketing* 56, no. 4: 1184–1209.

Ramírez-Martín, A., J. Ramos-Martín, F. Mayoral-Cleries, B. Moreno-Küstner, and J. Guzman-Parra. 2020. "Impulsivity, Decision-Making and Risk-Taking Behaviour in Bipolar Disorder: A Systematic Review and Meta-Analysis." *Psychological Medicine* 50, no. 13: 2141–2153.

Ratner, B. 2009. "The Correlation Coefficient: Its Values Range Between  $+1/-1$ , or Do They?" *Journal of Targeting, Measurement and Analysis for Marketing* 17, no. 2: 139–142.

Rogoza, R., C. M. Kowalski, and J. A. Schermer. 2019. "Dark Triad Traits Within the Framework of the Circumplex of Personality Meta-traits Model." *Journal of Individual Differences* 40, no. 3: 168–176.

Sahut, J. M., M. Laroche, and E. Braune. 2024. "Antecedents and Consequences of Fake Reviews in a Marketing Approach: An Overview and Synthesis." *Journal of Business Research* 175: 114572.

Saxena, A. 2018. "Deception Detection on Amazon Reviews Dataset." Github. <https://bit.ly/2W5Cw15>.

Spain, S. M., P. Harms, and J. M. LeBreton. 2014. "The Dark Side of Personality at Work." *Journal of Organizational Behavior* 35, no. S1: S41–S60.

Sumner, C., A. Byers, R. Boochever, and G. J. Park. 2012. "Predicting Dark Triad Personality Traits From Twitter Usage and a Linguistic Analysis of Tweets." In *2012 11th International Conference on Machine Learning and Applications* Vol. 2, 386–393. IEEE.

Trevino, L. K., and S. A. Youngblood. 1990. "Bad Apples in Bad Barrels: A Causal Analysis of Ethical Decision-Making Behavior." *Journal of Applied Psychology* 75, no. 4: 378–385.

UK Department for Business & Trade. 2023. *Fake Online Reviews Research: Estimating the Prevalence and Impact of Fake Online Reviews*. GOV.UK. <https://assets.publishing.service.gov.uk/media/6447c00c529eda000c3b03c5/fake-online-reviews-research.pdf>.

UK Government. 2024. *New Laws Set to Ban Mandatory Hidden Fees From Online Shopping, Saving Money for Consumers*. GOV.UK. <https://www.gov.uk/government/news/new-laws-set-to-ban-mandatory-hidden-fees-from-online-shopping-saving-money-for-consumers>.

Vallance, C. 2023. "Amazon, Expedia and Trustpilot Unite to Fight Fake Reviews." *BBC News*. <https://www.bbc.co.uk/news/technology-67156584>.

Walther, M., T. Jakobi, S. J. Watson, and G. Stevens. 2023. "A Systematic Literature Review About the Consumers' Side of Fake Review Detection—Which Cues Do Consumers Use to Determine the Veracity of Online User Reviews?" *Computers in Human Behavior Reports* 10: 100278.

Webb, T. L., and P. Sheeran. 2006. "Does Changing Behavioral Intentions Engender Behavior Change? A Meta-Analysis of the Experimental Evidence." *Psychological Bulletin* 132, no. 2: 249–268.

Wilson, D. S., D. Near, and R. R. Miller. 1996. "Machiavellianism: A Synthesis of the Evolutionary and Psychological Literatures." *Psychological Bulletin* 119, no. 2: 285–299.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Wu, Y., E. W. T. Ngai, P. Wu, and C. Wu. 2020. "Fake Online Reviews: Literature Review, Synthesis, and Directions for Future Research." *Decision Support Systems* 132: 113280.

Xiao, B., and I. Benbasat. 2011. "Product-Related Deception in e-Commerce: A Theoretical Perspective." *Mis Quarterly* 35: 169–195.

Yousaf, S., and J. M. Kim. 2023. "Dark Personalities and Online Reviews: A Textual Content Analysis of Review Generation, Consumption and Distribution." *Tourism Management* 98: 104771.

Zhang, D., L. Zhou, J. L. Kehoe, and I. Y. Kilic. 2016. "What Online Reviewer Behaviors Really Matter? Effects of Verbal and Nonverbal Behaviors on Detection of Fake Online Reviews." *Journal of Management Information Systems* 33, no. 2: 456–481.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.